

Propuesta preliminar de soporte para la integración de datos, metadatos y conocimiento geográfico mediante geo-ontologías

Rafael Oliva-Santos¹, Eduardo Garea-Llano², Francisco Maciá-Pérez³

¹Facultad de Matemática y Computación
Universidad de La Habana
roliva@matcom.uh.cu

²Centro de Aplicaciones de Tecnología de Avanzada
MINBAS, Ciudad de la Habana
egarea@cenatav.co.cu

³Departamento de Tecnología Informática y Computación,
Universidad de Alicante
pmacia@dtic.ua.es

Resumen. En este trabajo se presenta una primera propuesta a una nueva estructura ontológica con el objetivo de lograr la integración entre datos, metadatos y conocimiento geográfico. La principal contribución de la investigación consiste en la presentación de una nueva concepción de las geo-ontologías como estructuras de integración de conocimientos, datos y metadatos. Se presentan ejemplos de su aplicación para el tratamiento de la información espacial integrada a un Sistema de Información Geográfica en la temática del suelo y cobertura vegetal. El logro de una verdadera integración entre los datos geográficos, los metadatos correspondientes, y los conocimientos relacionados con ambos y la relación entre ellos es un reto en el momento actual para la creación de una Infraestructura de Datos Espaciales sobre una base semántica. El logro de esta integración contribuye a una mejor interoperabilidad entre datos y metadatos geográficos, así como a lograr una mejor administración en el análisis y la recuperación de la información que se lleva a cabo en los actuales sistemas computacionales y a través de la Web.

1 Introducción

El concepto de ontología ha atraído una atención creciente en la comunidad de las ciencias de la información debido a su capacidad para lograr una representación del conocimiento compartido. Las ontologías tienen una gran importancia en la creación y uso de las normas de intercambio de datos, así como en la solución de problemas derivados de la heterogeneidad y poca interoperabilidad de los datos geográficos. Las ontologías pueden ser usadas como una alternativa para representar los datos y, de forma explícita, el conocimiento acerca de ellos.

Los metadatos son datos acerca de los datos, describen características de grupos de datos y sus proveedores. El uso de metadatos permite una mejor recuperación de datos geográficos en la Web o en los Sistemas de Información Geográfica (SIG). A

excepción de los formatos de ficheros y bases de datos que incluyen implícitamente elementos de metadatos, la conexión de datos y metadatos es principalmente un acuerdo tomado por los usuarios. Por lo general, entre estas componentes no existe un enlace al mismo nivel que sea procesable automáticamente.

Una estructura de almacenamiento que permita la integración entre los datos y metadatos favorecería el intercambio de información geográfica. Esta estructura debe ser lo suficientemente flexible para que la variedad de fuentes y formatos de datos no constituyan un problema para su almacenamiento y gestión.

Los esfuerzos que se llevan a cabo por lograr la creación de Infraestructuras de Datos Espaciales (IDE) posibilitarán alcanzar la interoperabilidad entre datos, estándares de metadatos y conocimiento espacial. Para lograr ésto es necesario crear una forma de integración de la representación de datos y el conocimiento compartido entre las distintas comunidades y actores de esta infraestructura, así como los metadatos representados conforme a las normas internacionales.

En este trabajo, proponemos las geo-ontologías como estructuras en las que la integración de los datos, metadatos y conocimiento geográfico es posible. Presentamos una propuesta preliminar de la concepción de una estructura ontológica para la integración de estos tres componentes.

El resto del documento está organizado de la siguiente manera: en la sección 2 se describen los trabajos relacionados y algunos conceptos básicos necesarios; en la sección 3 se presenta la propuesta de geo-ontologías que integran los datos, metadatos y el conocimiento geográfico; la sección 4 muestra algunos ejemplos de los beneficios que se pueden obtener con nuestra propuesta, principalmente su integración a los SIG; en la sección 6 se ilustran dos posibles escenarios donde puede usarse la propuesta en la recuperación de la información de recursos naturales; finalmente las conclusiones y trabajos futuros se describen en la sección 6.

2 Trabajos relacionados y conceptos fundamentales

Cuando se menciona información espacial o geográfica en este trabajo, se refiere al conjunto de conceptos que la componen, especialmente los datos, metadatos y conocimiento. Los datos y metadatos geográfico tienen asociado conocimiento que generalmente no está estructurado de forma explícita. Gracias a la experiencia y los conocimientos acumulados, los especialistas pueden inferir conocimiento asociado a los datos y metadatos geográficos.

El uso de ontologías en la información geográfica tiende a ser diferente dependiendo de la perspectiva y objetivos de los usuarios [1]. Existen trabajos que proponen a las geo-ontologías como formas de representación del conocimiento [2], otros que las proponen como estructuras en las que los datos geográficos pueden ser codificados [3], otros como estructuras en las que se puede hacer persistentes los metadatos [4] y otros que las utilizan como elementos claves en la integración de datos de distintas fuentes [5]. En las siguientes secciones serán comentados los trabajos más relevantes en cada una de estas líneas de aplicación de las geo-ontologías, de esta manera se sentarán las bases para la propuesta de geo-ontologías como estructuras de integración de datos, metadatos y conocimiento geográfico.

2.1 Geo-ontologías como forma de representación del conocimiento

En el ámbito de la información espacial, los esfuerzos para gestionar el conocimiento se han centrado principalmente en el uso de ontologías y en la inclusión de elementos semánticos en datos geográficos. En los modelos de datos geográficos se representan un conjunto de los elementos básicos, propiedades y geometrías pero no todas las relaciones entre estos elementos están representadas en los modelos [3]. La falta de representación explícita del conocimiento inherente a la realidad instanciada en estos modelos afecta a la interoperabilidad semántica, por lo que se hace necesario incorporar representaciones del conocimiento, como las geo-ontologías, que complementen los datos y metadatos geográficos.

Gruber [6] define ontología como una especificación explícita de una conceptualización, Guarino [7] planteó la distinción entre una ontología y una conceptualización e inició el debate con la definición de conceptualización como un conjunto de relaciones conceptuales definidas en un espacio de dominio y la importancia del sentido y relevancia de estas relaciones con independencia de los elementos que las forman.

En este trabajo no consideramos las ontologías relevantes por poder ser utilizadas como un vocabulario compartido sino por el potencial semántico que le otorga el hecho de incluir axiomas y, de esta manera, ofrecer la posibilidad de emplear la lógica y poder inferir nuevos conocimientos. Las ontologías permiten hacer explícitos, identificar y asociar semánticamente conceptos relacionados [8], son representaciones formales del conocimiento.

Una geo-ontología es una ontología que ofrece una descripción de entidades geográficas y difiere de otras ontologías por la presencia predominante de relaciones topológicas que caracteriza a la información espacial. Una geo-ontología tiene dos tipos básicos de conceptos [9], los conceptos físicos, que son aquéllos que corresponden a fenómenos físicos en el mundo real y los conceptos sociales, que corresponden a las características del mundo que construimos social e institucionalmente. Según [10] y [9] los conceptos en una geo-ontología están dirigidos hacia objetos geográficos y éstos están organizados en una jerarquía donde el concepto raíz se denomina geo-objeto. El concepto raíz está especializado en dos clases que corresponden a los principales tipos de conceptos geográficos: los relacionados con el fenómeno de la continuidad y objetos individuales.

El uso de las geo-ontologías como formas de representación del conocimiento constituye un paso de avance en la solución de los problemas derivados de la heterogeneidad semántica en los datos geográficos. Las geo-ontologías pueden utilizarse para hacer explícita la semántica del contenido de servicios Web, con el fin de mejorar el descubrimiento y recuperación de la información geográfica en el entorno Web [2]. Se ha demostrado la utilidad de la integración de mecanismos de descubrimientos basados en ontologías a sistemas de información, lo que posibilita superar problemas derivados de la heterogeneidad semántica [2].

En [2] muestra como las geo-ontologías pueden ser soportes para realizar conexiones semánticas durante los procesos de búsqueda y recuperación de información espacial. El uso de geo-ontologías para describir los geo-servicios favorece la interoperabilidad entre ellos y permite su integración [11].

2.2 Geo-ontologías, estructuras que son capaces de almacenar datos

La información geográfica se caracteriza por la capacidad de integrar conjuntos de información, que de otra forma resultan disjuntos, mediante la aplicación de las relaciones espaciales de coincidencia, proximidad o adyacencia. Ésta es la característica que proporciona mayor potencialidad a la información geográfica y que constituye la base del análisis espacial. Cada vez son más frecuentes los fenómenos que tienen, en su descripción, alguna componente espacial subyacente que es la determinante para mejorar la comprensión de su dinámica de funcionamiento [12].

Con frecuencia, los análisis espaciales requieren una descripción precisa de los objetos geográficos para obtener buenos resultados, sin embargo la imprecisión de las fuentes son diversas: diferentes escalas, los diferentes niveles de resolución, o atributos que están implícitos en la composición geográfica de los objetos. Merece la pena analizar alternativas de representación conceptual que son independientes de la precisión de los datos.

En [13] se obtuvo un *framework* que permite la correspondencia entre una geo-ontología y un esquema conceptual geográfico. De esta manera se logra vincular la representación formal de la semántica para describir esquemas conceptuales y la información almacenada en bases de datos.

Las ontologías son una alternativa para la representación de datos geográficos. En [3] se propusieron las ontologías para describir y representar datos geográficos vectoriales. La representación propuesta permite la manipulación de datos imprecisos, ya que traduce en un esquema conceptual la descripción de un mapa cartográfico.

En la mayoría de los artículos sobre geo-ontologías, las utilizan como conceptualizaciones generales, conceptos que pueden ser compartidos, lo que permite que sean muy importantes para lograr una mejor interoperabilidad. En este trabajo, no se refiere a las geo-ontologías como conceptualizaciones generales, sino como representaciones formales de conocimiento y representaciones conceptuales de los datos geográficos.

Un ejemplo de la utilización de ontologías en la conceptualización de elementos geográficos son las ontologías desarrollados en OWL por [14] que representa la superficie y los modelos numéricos de calidad de aguas basado en el *Surface Water and Water Quality Models Information Clearinghouse* de *US Geological Survey (USGS)*.

2.3 Geo-ontologías, estructuras que son capaces de almacenar metadatos

Los metadatos estructuran el contenido de los datos, información sobre ellos y contienen algunos elementos semánticos. En general, no están estructuralmente preparados para representar el conocimiento. En las ontologías se puede almacenar información como la que normalmente se almacena en metadatos y como son formas de representación del conocimiento permiten la estructuración de la semántica asociada a esta información.

Los metadatos de información geográfica constituyen uno de los primeros pasos en la administración de los elementos semánticos de los datos geográficos. En las normas actuales no es posible representar los conocimientos asociados a la información

La utilización de ontologías tiene como primer paso la creación de un vocabulario para el intercambio de conocimientos en el ámbito en el que se enmarca la ontología. Ésta es una de las líneas a seguir en el desarrollo de estándares de metadatos. Las ontologías son estructuras en las que es posible representar los metadatos y los conocimientos relacionados con ellos, permiten diferenciar entre el conocimiento del dominio y el conocimiento operacional. En ontologías, también es posible separar estructuralmente las técnicas de los algoritmos.

Las geo-ontologías permiten la codificación de diferentes normas de metadatos, o una fusión de las mismas, para que éstos se puedan recuperar en función de alguna de las normas que se fusionaron. Las dos normas internacionales de metadatos más importantes son: el estándar estadounidense *Federal Geographic Committee Dates (FGDC-STD-001-1998)* de *Content Standard Digital for Geospatial Metadata (CSDGM)* versión 2 y la norma ISO/TC 211 19115-2003 de *International Standard Organization (ISO)*. Esta norma ISO para metadatos se expresa mediante diagramas UML (*Unified Model Language*) que pueden ser representados textualmente en ficheros XML (*Extensible Markup Language*). La comunidad de desarrolladores de SIG está trabajando en la inclusión de los elementos semánticos en representaciones de estas normas y estándares, en [15] se desarrolló una geo-ontología de dominio tomando como punto de partida los diagramas UML que esboza la norma ISO 19115.

El modelado de metadatos en estructuras ontológicas permite hacer persistentes los conocimientos asociados a los elementos que conforman los metadatos. Con la aparición del concepto de Web Semántica, se comenzó a desarrollar ontologías que representan semánticamente metadatos geográficos, ejemplo de esto es la geo-ontología de Hidrología y Administración desarrollada por el *Ordnance Survey* del Reino Unido. En *Drexel University* se han desarrollado varias ontologías que modelan las principales normas de metadatos geográficos, como son las ontología en OWL DL para los metadatos de acuerdo a la norma ISO 19115 2003 [4] y la norma CSDGM del *Federal Geographic Data Committee (FGDC)* [16].

2.4 Heterogeneidad en los datos geográficos

La heterogeneidad de los datos geográficos es un problema que limita la mayoría de los procesos que utilizan este tipo de información. En [11] se esbozan tres tipos de heterogeneidad en los datos geográficos: heterogeneidad sintáctica, heterogeneidad estructural y heterogeneidad semántica.

La heterogeneidad sintáctica radica en los distintos modos de codificar los datos. Hay diferentes formatos de datos, los datos geográficos pueden ser representados principalmente en dos modelos (vectorial o matricial (*raster*)), y pueden referirse a diferentes sistemas de coordenadas espaciales.

La heterogeneidad estructural radica en los diferentes métodos utilizados para organizar información. Un elemento importante para evitar la heterogeneidad estructural es el uso de metadatos porque describen la estructura de la representación del esquema en los datos.

La heterogeneidad semántica se refiere a los diferentes usos de un mismo modelo de datos por los agentes (personas, organizaciones o sistemas computacionales), que está dado por la no codificación explícita del conocimiento relativo al modelo de

datos, como puede ser utilizar diferentes terminologías o simbología para referirse a la misma información.

La heterogeneidad de los datos geográficos es un tema que ha sido tratado [17] y en el que en la actualidad se exigen soluciones.

2.5 Geo-ontologías, estructuras utilizadas en la integración de datos geográficos

Muchas investigaciones están encaminadas a encontrar la manera de codificar formalmente el contexto y las relaciones geográficas. Un tipo de relación de este ámbito son las que se existen entre datos y metadatos geográficos, así como con el conocimiento entorno a estos y su interrelación.

La integración de datos y metadatos ha evolucionado partiendo de la simple enumeración de los metadatos asociados a los datos para su gestión en un SIG. La mayoría de los SIG y los más populares formatos de datos (ESRI, MAPINFO, Ilwis, y ERDAS) incluyen algunos elementos de las normas de metadatos. Los SIG actuales dan soporte a la creación y edición de metadatos. En este sentido, Jack Dangermond, presidente de ESRI, considera que las herramientas y sistemas para la creación, edición y gestión de los metadatos debe ser una característica estándar incorporada a cualquier SIG [18]. No se conoce un formato de datos comerciales que esté completamente integrado con las normas de metadatos. Los formatos actuales de datos no son lo suficientemente flexibles para propiciar la integración con alguna norma de metadatos.

Sobre la integración de fuentes de datos geográficos sí se tienen resultados considerables. Existen varios trabajos que comparan y evalúan las propuestas de integración de datos geográficos. En [5] se introduce el sistema BUSTER (*Bremen University Semantic Translator for Enhanced Retrieval*) que propone una arquitectura formada por dos subsistemas: el primero tiene como objetivo resolver los problemas de recuperación de información mediante búsquedas inteligentes y el segundo con el objetivo de resolver la heterogeneidad semántica propone una forma de integración de la información. En [19] se presenta una metodología para explorar e identificar información semántica a partir de una categorización provista por geo-ontologías, esta metodología propone medidas sobre qué ontologías pueden ser integradas o asociadas mediante categorías.

En [20] se propone una metodología para la integración de esquemas conceptuales espacio-temporales mediante modelos conceptuales y Lógica Descriptiva (LD). En [21] se introduce una arquitectura y una metodología basada en LD para crear un sistema integrado de información geográfica. Las ontologías fueron propuestas en [22] como mediadoras en la integración semántica de datos geográficos.

Un ejemplo de integración de información geográfica a nivel de sistema son los SIG gobernado por ontologías (SIGGO – del inglés *ODGIS Ontology Driven Geographic Information System*) que actúan como un sistema integrador independientemente del modelo [23]. En el sistema GioNis [24] se define una propuesta de ontología híbrida basada en una arquitectura semántica que combinada con LD permite descubrir correspondencias entre conceptos de diferentes ontologías. En [25] describe una metodología para desarrollar un SIG integrado principalmente

para resolver los problemas de heterogeneidad semántica en bases de datos topográficas.

2.6 Consideraciones sobre la integración de datos, metadatos y conocimiento geográfico

En cuanto a integración en el ámbito geográfico, como se ha visto en la sección anterior, los principales trabajos están dirigidos a la integración de distintas fuentes de datos geográficos. Los sistemas actuales separan los metadatos del proceso de integración y análisis de los datos geográficos lo que limita la interoperabilidad [26]. La integración de datos y metadatos geográficos es necesaria para aumentar la interoperabilidad entre sistemas, disminuir la heterogeneidad entre los datos y servicios y aprovechar en los SIG al máximo los recursos disponibles.

Las ontologías son estructuras en esta integración es posible. Hay varios trabajos que las proponen para representar conocimiento geográfico [2, 9, 10], otros para representar metadatos [4, 15, 16] y otros datos [3, 14]. No se conocen trabajos sobre la integración de estos tres componentes. Como hay trabajos que demuestran que cada componente puede formalizarse en ontologías, la interrelación entre ellos también puede formalizarse mediante recursos como relaciones, funciones, propiedades y axiomas.

El uso de ontologías en procesos de integración ha dado buenos resultados. En [2] se plantea que la integración a SIG de mecanismos de descubrimientos basados en ontologías elimina problemas derivados de la heterogeneidad semántica. En [21] se propone el uso de ontologías para resolver la heterogeneidad semántica en la integración de Bases de Datos Espaciales.

3 Propuesta de estructura ontológica para la integración de los datos, metadatos y el conocimiento

En la actualidad, no existe una unidad semántica en la información espacial. Diferentes organizaciones grafican en el plano, de manera distinta, instancias de un mismo concepto, no hay un acuerdo sobre el significado de la representación de la información espacial. Por ejemplo las líneas pueden representar para una organización *autopistas*, mientras que para otra *ríos*. El uso de conceptualizaciones compartidas permite que en diferentes mapas o bases cartográficas se pueda representar de la misma manera un fenómeno dado. La unidad semántica de los datos geográficos favorece la interoperabilidad y el intercambio de datos entre los diferentes sistemas computacionales.

3.1 Geo-ontologías, estructuras que facilitan la integración de los datos, metadatos y el conocimiento

La vinculación entre los datos y metadatos es generalmente un acuerdo adoptado por los creadores de los datos. No se conocen formatos de datos que permitan establecer una conexión con metadatos que sigan las principales normas internacionales que pueda ser procesable automáticamente, lo que afecta la organización y uso adecuado de los datos geográficos. Las ontologías son estructuras suficientemente adaptables en las que la integración de las diferentes componentes de la información espacial (dato, metadato y conocimiento geográfico) es posible.

El uso de ontologías permite hacer explícita la representación del conocimiento relacionado con los datos geográficos y metadatos que antes debía ser inferido por los especialistas. La explotación de datos y metadatos guiado por el conocimiento almacenado en ontologías son más apropiados y orientados a los intereses con los que fueron creados.

En este trabajo proponemos las geo-ontologías como las estructuras que garantizan la viabilidad de gestión automática y la persistencia de la integración entre los datos, metadatos y el conocimiento geográfico. En la Fig. 1 se muestran estos tres grandes componentes de la información espacial y sus vínculos. La intersección entre los formatos de datos espaciales y metadatos no es significativa, y el vínculo entre estos dos componentes está dado principalmente a nivel de SIG con recursos ajenos al dato y metadato, se muestra con una flecha bidireccional más gruesa que las demás pero discontinua porque la integración no es a nivel de dato y metadato. El conocimiento geográfico sobre el dato, metadato y su vinculación no está explícitamente persistente en la mayoría de los caso de modo que pueda ser procesable automáticamente, se muestra con flechas bidireccionales más delgadas con el objetivo de mostrar el vínculo pero en este caso es prácticamente inexistente desde el punto de vista computacional.

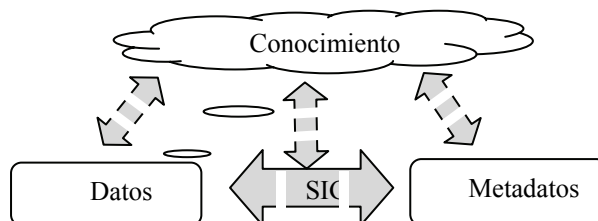


Fig. 1. Vínculos actuales entre los componentes de la información espacial.

En el caso del conocimiento geográfico no se encontraron trabajos significativos sobre la persistencia de conocimientos entorno al dato, metadato o su interrelación. En este sentido la mayor parte de los trabajos están orientados en la creación y diseño de ontologías de dominio geográfico [27].

La propuesta preliminar de estructura ontológica de integración de datos, metadatos y conocimiento geográfico se ilustra en el esquema de la Fig. 2. En este caso la geo-ontología será la estructura base donde se llevará a cabo la integración entre datos y metadatos, esta puede ser física en el caso que ambos componentes se codifiquen en la propia geo-ontología o mediante referencias en caso que alguno de estos componentes esté disponible en un servidor o ubicación física distinta. Sobre

esta relación se pueden incluir en la estructura conocimiento que guíe su utilización y den detalles de importancia a tener en cuenta en su recuperación. De la misma manera la estructura facilita la inclusión de descripciones semánticas de datos y metadatos geográficos.

En este caso los vínculos son mostrados con flechas bidireccionales continuas y de relleno sólido puesto que son procesables computacionalmente y serán persistentes en la estructura propuesta. Además, la vinculación dato-metadato será tratada como un nuevo componente el cual tendrá asociado conocimiento que también será persistente en la estructura.

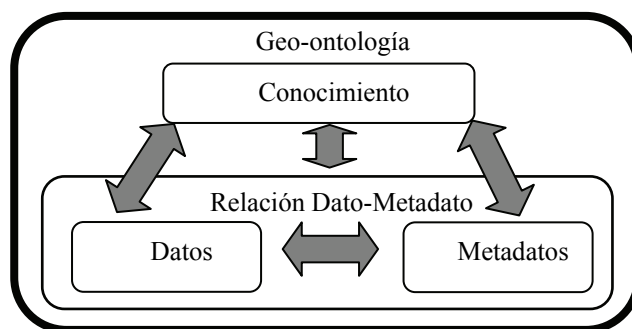


Fig. 2. Diagrama de propuesta de estructura de integración entre datos, metadatos y conocimiento geográfico.

Utilizar a las geo-ontologías como estructuras de integración de datos, metadatos y conocimiento geográfico permite poder explotar al máximo todas las potencialidades de las ontologías como mecanismo de representación formal del conocimiento, como son inclusión de axiomas, nuevos conceptos, reutilización del conocimiento previamente formalizado en otras ontologías, así como la capacidad de integrarse con otras ontologías y ser reutilizadas en la construcción de ontologías de otros ámbitos.

Con esta propuesta no se niega la importancia y la eficiencia del almacenamiento de información espacial en múltiples archivos, ni la estructuración de los metadatos en jerarquías. En estos casos, nuestra propuesta de estructura de integración puede estar dividida físicamente pero conectada semánticamente de manera que continúe siendo integradora. Es imprescindible que la recuperación de cualquier componente partiendo de otro componente o de información o conocimiento suministrado sea eficiente y eficaz. Esta estructura necesita un mecanismo que debe ser capaz de recuperar los datos, metadatos y de los conocimientos como un todo, aunque físicamente se almacenan separados. La propuesta de estructura de integración debe contener tanto la formalización del conocimiento sobre los componentes como de sus interrelaciones, además de cuestiones de almacenamiento como la ubicación física donde están almacenados en caso de no estar físicamente codificados en la propia estructura.

La utilización de geo-ontologías como estructura de integración nos permite tener una estructura única para la gestión de todos los componentes de la información espacial y constituye un paso de avance a favor de la unidad semántica espacial. La incorporación de datos geográficos a geo-ontologías posibilita una mayor integración

datos-conocimientos y facilita la interoperabilidad semántica entre los sistemas que manejan información espacial.

4 Beneficios que se pueden obtener con la estructura de integración propuesta

Las ontologías permiten la integración de los servicios relacionados con la información espacial [11]. La utilización de ontologías para describir los geo-servicios favorece la interoperabilidad entre ellos. La representación en la misma estructura de datos, metadatos y conocimientos geográficos permite un mejor desempeño de los geo-servicios. Nuestra propuesta de estructura de integración beneficia el intercambio de datos, ya que reúne en un recurso computacionalmente gestionable datos, metadatos y conocimiento geográfico.

Entre los beneficios más importantes de nuestra propuesta podemos mencionar: en primer lugar, la importancia de su integración en los SIG, en segundo lugar su importancia para evitar la heterogeneidad y por último, las posibilidades que tiene para la obtención de perfiles y vistas dinámicas de metadatos.

4.1 Geo-ontologías, estructuras de integración en los Sistemas de Información Geográfica Gobernado por Ontologías

La diversidad de formatos de datos geográficos, así como las distintas conceptualizaciones empleadas para la descripción y explotación de los datos es un obstáculo para la interoperabilidad entre los SIG. En este sentido, existe un interés marcado por parte de la mayoría de los gobiernos en el desarrollo de Infraestructuras de Datos Espaciales (IDE) [28] con el objetivo de establecer y definir las políticas, tecnologías, estándares y recursos humanos necesarios para la efectiva recolección, procesamiento de datos, administración, acceso, entrega y utilización de los datos geográficos en los distintos niveles regionales en función de las decisiones económicas, políticas, sociales y del desarrollo sostenible. La interoperabilidad entre estas IDE será posible en la medida que exista un basamento conceptual compartido. Este interés internacional enfatiza la necesidad de lograr la unidad semántica entre los datos espaciales.

Los SIG constituyen la herramienta fundamental para el desarrollo de aplicaciones en la gestión y análisis de información espacial integrada de las IDE. El uso de ontologías en la construcción de SIG comienza a aparecer en el 1997 en los trabajos de Frank [28]. En [23] se plantea que el primer paso para construir la nueva generación de SIG sería la construcción de una colección sistemática y específica de entidades geográficas, sus propiedades y relaciones. En años recientes, en [23] se introdujo una nueva arquitectura SIG que puede facilitar la integración de manera flexible y se basa en el valor semántico de la información, sin tener en cuenta su representación, los SIGGO.

Un SIGGO evita la clasificación de los datos sobre la base de sus representaciones solamente. El enfoque semántico, basado en el concepto de entidades geográficas

[23], posibilita una integración natural de diferentes tipos de información a través del uso de clases flexibles. El primer paso para construir una SIGGO es la especificación de ontologías por medio de un editor de ontologías. El editor almacena una representación formal de ontologías y proporciona una traducción a un lenguaje de programación de alto nivel (por ejemplo Java). Después de la traducción, las ontologías estarán disponibles en forma de clases.

En la Fig. 3, se muestra una posible incorporación de la propuesta de estructura de integración en los SIGGO. Nuestra propuesta, enmarcada en un círculo con un grosor de línea mayor, sustituiría a los datos en el esquema original planteado en [23]. De este modo el uso y acceso a los datos estará influenciado por los metadatos correspondientes así como por el conocimiento específico sobre él.

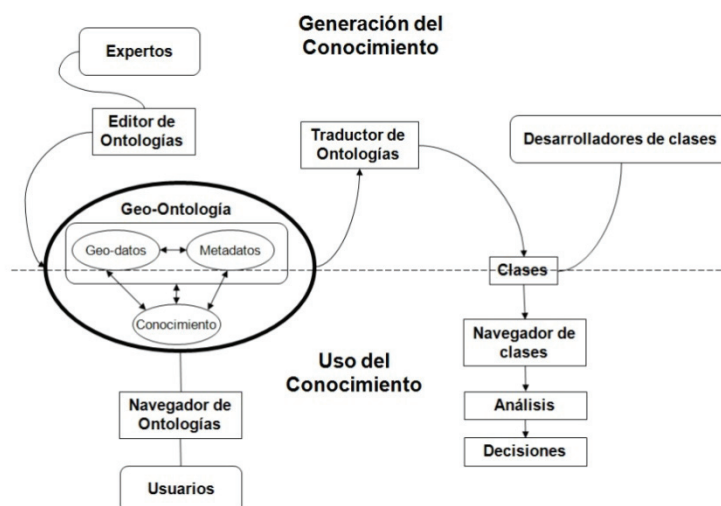


Fig. 3. Incorporación de la propuesta de estructura de integración al diagrama de SIGGO propuesto por sus creadores [23].

4.2 Posible solución a la heterogeneidad de la información espacial

Nuestra propuesta de estructura integradora ayuda a resolver el problema de la gran heterogeneidad presente en la información espacial. En [29] mediante el uso de ontologías se solucionan problemas derivados de la heterogeneidad semántica entre documentos GML (*Geography Markup Language*). Un paso importante para poder resolver la heterogeneidad será la definición de normas que especifiquen la manera de llevar a cabo la integración en este tipo de estructura, principalmente desde el punto de vista de la semántica.

La utilización de ontologías para representar datos geográficos permite que sean representados distintos modelos de datos lo que permite evitar la heterogeneidad sintáctica. El GML y los lenguajes más populares de definición de ontologías para la naciente Web Semántica como OWL (*Web Ontology Language*) y RDFS (*Resource Descriptor Framework Schema*), tienen su raíz en el XML. En [3] se conceptualiza

información espacial en OWL. El uso de solo una manera de representar la información espacial hace que no haya cabida para la heterogeneidad estructural, ya que será la misma estructura ontológica en la que se representará todos los tipos de datos y metadatos geográficos.

En la propuesta de estructura integradora no tiene lugar la heterogeneidad semántica ya que se basa en ontologías y éstas son representaciones formales del conocimiento y estructuralmente son capaces de almacenar conceptos, relaciones, propiedades, instancias de los conceptos y axiomas. La semántica de los datos y metadatos geográficos, así como de su relación será almacenada o referenciada en la misma estructura, lo que permite que la correcta explotación de cada uno de estos componentes pueda estar guiada por el conocimiento formalizado en la estructura propuesta. La inclusión de elementos semánticos en la representación espacial permite la interoperabilidad entre sistemas de una forma completamente automática sin intervención humana y que abre el camino a la incorporación de los SIG con la Web Semántica.

4.3 Construcción de perfiles de las normas de metadatos geográficos

Durante los últimos años, diferentes organizaciones y gobiernos han llevado a cabo un arduo trabajo de elaboración de normas y estándares de metadatos geográficos. La tarea de armonizar la gama de las normas oficiales durante el período 1999-2002 estuvo encabezada por el ISO / TC 211 y como resultado se creó en el 2003 la norma ISO 19115. Las distintas comunidades adecuan las normas internacionales a sus situaciones concretas mediante el trabajo con perfiles o subconjuntos de las normas, incluso en ocasiones con la inclusión de otros elementos en calidad de ampliación de la norma.

Las ontologías son un mecanismo de estructuración que permite la integración de las normas, así como la selección de ciertos elementos para la formación de perfiles y sistemas dinámicos de gestión de metadatos. Debido a la flexibilidad de la propuesta de estructura de integración dada su base ontológica, es posible obtener diferentes perfiles de los metadatos almacenados, además que la estructura de integración propuesta permite la codificación de metadatos según la integración de varias normas.

Las normas estructuran y normalizan la información mediante el establecimiento de niveles de obligatoriedad e importancia sobre la información que se almacena, pero no incluyen la semántica asociada a esta información. El conocimiento inherente a las relaciones entre los elementos que componen la norma no se hace persistente en ella. La semántica en torno a los metadatos no es persistente en las normas y se deja abierto a la interpretación de los usuarios finales.

Los elementos relacionados con la semántica de los datos y metadatos no están sujetos a la normalización aún. Nuestra propuesta refuerza la necesidad de la normalización de nuevos elementos que serán almacenados explícitamente en la estructura de integración, como son el conocimiento en torno a los datos, metadatos y su interrelación. En estos momentos un grupo de expertos de la ISO 19150 está explorando cómo los estándares ISO 19100 pueden beneficiar los objetivos de ISO/TC 211 y contribuir a la Web semántica [30].

5 Dos posibles escenarios de la utilización de las geo-ontologías como estructuras de integración

En esta sección presentamos dos posibles escenarios donde es posible la utilización de esta primera aproximación a la estructura ontológica para llevar a cabo operaciones de recuperación de información en un SIG que integra información para la gestión de los recursos naturales en las regiones montañosas. La información geográfica utilizada corresponde a una región montañosa de la parte oriental de Cuba y se almacena en dos mapas vectoriales. Los mapas en la Fig. 4 se muestran las dos capas temáticas utilizadas. Cada capa contiene objetos geográficos representados por primitivas espaciales (polígonos). Estos mapas muestran los suelos y la cubierta vegetal.

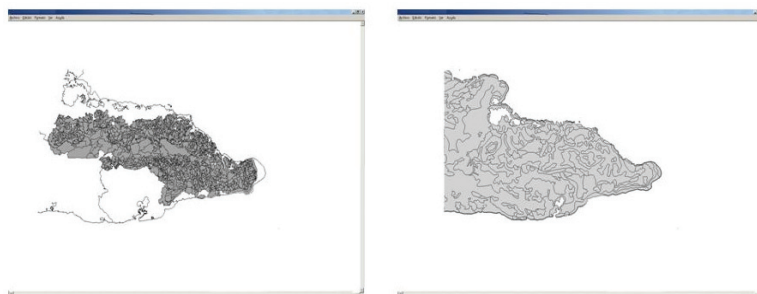


Fig. 4. Capas temáticas utilizadas en el análisis a) suelos b) cobertura vegetal.

Escenario 1: Recuperar la información del mapa de suelos siguiendo los siguientes criterios de decisión dados por un experto: determinar la distribución espacial del suelo *Ferralsol Rhodic* con las siguientes características: profundidad pedológica mayor de 30 cm y el contenido de materia orgánica superior a 10%.

En primer lugar, el usuario revisa el servidor de ontologías en busca de clases relacionadas con objetos de tipo polígono que estén contenidos en la capa de suelo. A continuación, el servidor de ontologías comienza la búsqueda de información que haga referencia a los atributos con los valores requeridos, en este caso: el tipo de suelo (TS = "*Ferralsol*"), profundidad pedológica (PD > 30,0 cm) y el contenido de materia orgánica (MO > 10%) y los metadatos asociados a ellos. Luego, devuelve una serie de objetos de la clase específica que cumple con todas las condiciones indicadas. A continuación se muestra la distribución espacial de los objetos seleccionados en el mapa (Fig. 5).

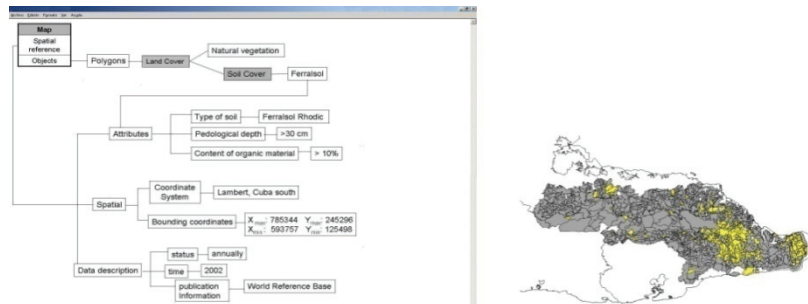


Fig. 5. Resultado de la recuperación de la información sobre los suelos en el SIGGO.

Escenario 2: Recuperación de información del mapa de cubierta vegetal siguiendo los siguientes criterios de decisión dados por un experto: determinar qué objetos geográficos de tipo 'selva tropical' tienen densidad de árboles mayor de 3 por m^2 y altura media de la selva mayor de 5m. Similar al caso anterior, el usuario revisa el servidor de ontología en busca de clases relacionadas con los objetos de tipo polígono contenidos en la capa de cobertura vegetal. A continuación, el servidor de ontologías comienza la búsqueda de información que haga referencia a los atributos con los valores requeridos, en este caso: el tipo de vegetación (TB = "selva tropical"), de densidad de plantación ($DA > 3$ árboles/ m^2), la vegetación de altura ($Vh > 5m$) y los metadatos asociados a ellos. Luego del análisis se devuelven una serie de objetos de que cumplen con las condiciones indicadas y se pasa a mostrar la distribución espacial de los objetos seleccionados en el mapa (Fig. 6).

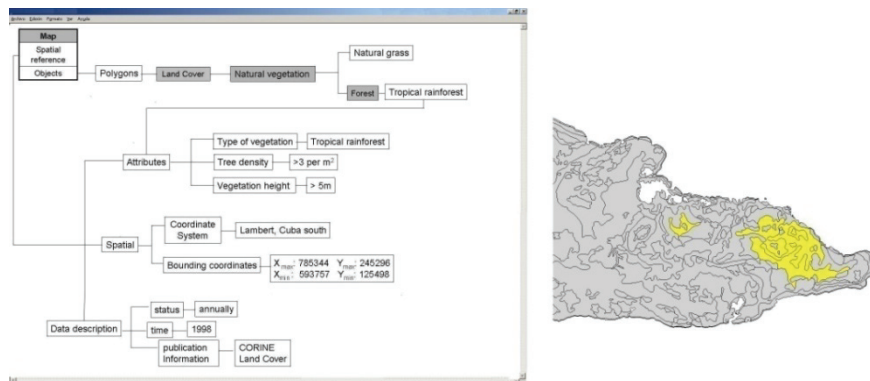


Fig.6. Resultado en la recuperación de información de la cobertura vegetal en un SIGGO.

6 Conclusiones y trabajos futuros

El principal resultado de esta investigación es la presentación de las ontologías como estructuras que permiten la integración de datos geográficos con sus metadatos y con el conocimiento asociado a los dos y su interrelación. Esta nueva propuesta de

esquema de información espacial permite la unificación de sus componentes y obtiene todos los beneficios que aportan la gestión y administración en los SIG de una sola fuente de información integrada.

La propuesta de estructura de integración favorece la integración de la información geográfica en la Web Semántica y constituye un paso adelante en la solución del problema de la heterogeneidad de la información geográfica. El uso de este tipo de esquema espacial permite la unificación de criterios acerca de los datos, metadatos y conocimiento espacial y da un nuevo sentido a la normalización de la información espacial haciendo énfasis en la necesidad de elaborar normas que regulen cómo hacer persistente explícitamente la semántica espacial en formas de representación del conocimiento como las ontologías.

Otros de los beneficios de integrar en ontologías, datos, metadatos y conocimiento geográfico, que se ilustran en este trabajo son: la obtención dinámica de perfiles de metadatos y la incorporación de la estructura ontológica a los SIGGO. La utilización de ontologías como representaciones conceptuales en los SIG permite la gestión de elementos semánticos, facilitando el desarrollo de tareas como la búsqueda en mapas interactivos y constituye la base para una verdadera interoperabilidad basada en el conocimiento.

Entre las líneas futuras está trabajar en la implementación de esta primera idea de estructura de integración, así como el estudio y la especialización de la propuesta para su adaptación a los sistemas multificheros que interactúen con jerarquías de metadatos para lograr una administración eficaz de la información espacial.

Referencias

1. Winter, S., Ontology: Buzzword or paradigm shift in GI science? *International Journal of Geographical Information Science*, 2001. 15(7): p. 587-590.
2. Lutz, M. and E. Klien, Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science*, 2006. 20(3): p. 233-260.
3. Torres, M., et al., Ontology-driven description of spatial data for their semantic processing, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2005: Mexico City. p. 242-249.
4. Saiful-Islam, A., et al. Ontology for Geographic Information - Metadata (ISO 19115:2003). 2004 [cited 2007-05-05]; Available from: <http://loki.cae.drexel.edu/~wbs/ontology/2004/09/iso-19115.owl>
5. Visser, U., *Intelligent Information Integration for the Semantic Web*. LNAI, 2005. 3159.
6. Gruber, T.R., A translation approach to portable ontology specifications. *Knowledge Acquisition*, 1993. 5(2): p. 199-220.
7. Guarino, N., *Formal Ontology in Information Systems*. IOS Press, Amsterdam (1998). Amended version of previous one in *Proceedings of the 1st International Conference*, 1998.
8. Wache, H., et al., Ontology-based integration of information - A survey of existing approaches. *IJCAI-01 Workshop: Ontologies and Information Sharing*, 2001: p. 108-117.
9. Fonseca, F. and G. CÃğmara, A Framework for Measuring the Interoperability of Geo-Ontology. *Spatial Cognition and Computation*, 2006. 6: p. 307-329.

10. Smith, B. and D.M. Mark, Ontology and geographic kinds. Proc. Int. Symposium on Spatial Data Handling, SDH'98, 2004: p. 1998.
11. Lemmens, R., Semantic Interoperability of Distributed Geo-services. Delft, Netherlands. NCG, 2006: p. 76.
12. Longley, P.A., et al., GIS Data Collection, Geographic Information Systems and Science, 2001: p. 32-37.
13. Fonseca, F., C. Davis, and C. Câmara, Bridging ontologies and conceptual schema in geographical information integration. *GeoInformatica*, 2003. 7(4): p. 307-321.
14. Saiful-Islam, A. and M. Piasecki. Ontology for Surface Water and Water Quality Models. 2004 [cited 2007-05-05]; Available from: <http://loki.cae.drexel.edu/~wbs/ontology/2004/04/model.owl>.
15. Probst, F., F. Gibotti, and A. Pazos, Connecting ISO and OGC models to the semantic web. Proceedings of the Third International Conference on Geographic Information Science, 2004: p. 181-184.
16. Saiful-Islam, A., et al. Ontology for Content Standard for Digital Geospatial Metadata (CSDGM) of Federal Geographic Data Committee (FGDC). 2004 [cited 2007-05-05]; Available from: <http://loki.cae.drexel.edu/~wbs/ontology/2004/08/fgdc-csdgm.owl>
17. Lutz, M., et al., Overcoming semantic heterogeneity in spatial data infrastructures. *Computers and Geosciences*, 2009. 35(4): p. 739-752.
18. ESRI, Metadata and GIS. 2002.
19. Kavouras, M., M. Kokla, and E. Tomai, Comparing categories among geographic ontologies. *Computers and Geosciences*, 2005. 31(2): p. 145-154.
20. Sotnykova, A., et al., Semantic mappings in description logics for spatio-temporal database schema integration, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, E. Zimanyi, Editor. 2005. p. 143-167.
21. Hakimpour, F. and A. Geppert, Resolution of semantic heterogeneity in database schema integration using formal ontologies. *Information Technology and Management*, 2005. 6(1): p. 97-122.
22. Hess, G.N. and C. Iochpe, Ontology-driven resolution of semantic heterogeneities in gdb conceptual schemas. Proceedings of the GEOINFO'04: VI Brazilian Symposium on GeoInformatics, 2004: p. 247-263.
23. Fonseca, F., Ontology-driven geographic information systems. *Ontology-driven Geographic Information Systems*, 2001.
24. Stoimenov, L., A. Stanimirovic, and S. Djordjevic-Kajan, Discovering mappings between ontologies in semantic integration process. Proceedings of the AGILE 2006, Visegrád, 2006: p. 213-219.
25. Aerts, K., K. Maesen, and A. van Rompaey, A practical example of semantic interoperability of large-scale topographic databases using semantic web technologies[C]. Proceedings of the AGILE'06, Visegrád, Hungary (2006) 35-42, 2009.
26. Schuurman, N., Formalization matters: Critical GIS and ontology research. *Annals of the Association of American Geographers*, 2006. 96(4): p. 726-739.
27. Balmaseda, C., Construyendo ontologías para el dominio de Suelos, in *V Congreso Internacional de Geomática*. 2007: La Habana (Cuba).
28. Frank, A.U., Spatial ontology: A geographical information point of view. *Spatial and Temporal Reasoning*, 1997: p. 135-153.
29. Córcoles, J.E., P. González, and V. López Jaquero, Integration of spatial XML documents with RDF, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2003. p. 407-410.
30. Cooper, A. and D. Danko, Actividades ISO/TC 211 in ICA News. 2006. p. 12.